DOCUMENT RESUME

ED 363 648 TM 020 695

AUTHOR	Blankmeyer, Eric
TITLE	Orthogonal Regression and Equivariance.
PUB DATE	93
NOTE	12p.
PUB TYPE	Statistical Data (110) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Equations (Mathematics); *Estimation (Mathematics); *Least Squares Statistics; Mathematical Models; *Regression (Statistics); *Research Methodology; *Robustness (Statistics)
IDENTIFIERS	*Equivariance; *Orthogonal Comparison

ABSTRACT

Ordinary least-squares regression treats the variables asymmetrically, designating a dependent variable and one or more independent variables. When it is not obvious how to make this distinction, a researcher may prefer to use orthogonal regression, which treats the variables symmetrically. However, the usual procedure for orthogonal regression is not equivariant. A simple modification is proposed to overcome this serious defect. Illustrative computations involving 15 observations on 5 variables are provided, and a robust version of the method is discussed. The modified orthogonal regression allows a researcher to explore a symmetric, equivariant, and robust linear relationship among a set of variables. (Contains 6 references.) (Author/SLD)

Orthogonal Regression and Equivariance

Eric Blankmeyer

Department of Finance and Economics

Southwest Texas State University

San Marcos, TX 78666

Tel. 512-245-2547

Abstract. Ordinary least-squares regression treats the variables asymmetrically, designating a dependent variable and one or more independent variables. When it is not obvious how to make this distinction, a researcher may prefer to use orthogonal regression, which treats the variables symmetrically. However, the usual procedure for orthogonal regression is not equivariant. We propose a simple modification to overcome this serious defect. Illustrative computations are provided, and a robust version of our method is discussed.

Key words: least squares regression, orthogonal regression, equivariance, robust estimation.

Orthogonal Regression and Equivariance

1. Introduction

To use ordinary least squares, one designates a dependent variable and one or more independent variables. This decision implies that the random error affects <u>only</u> the dependent variable. The choice of the dependent variable will usually be crucial for parameter estimates and the outcome of hypothesis tests. Sometimes considerations of cause and effect make it clear which variable is dependent and which are independent. Often, however, a researcher has no such preconception and prefers to treat the variables symmetrically.

In that case, each variable is equally subject to the random error. An appropriate linear model is orthogonal regression, where the error is not measured along one axis. Instead it is measured perpendicular to the regression plane itself, the usual Euclidean notion of the distance from a point to a line [Morrison (1990), chapter 8].

Despite its appealing symmetry, this method has a major disadvantage: the coefficients in an orthogonal regression are <u>not equivariant</u>; they change in a complicated way when a variable is rescaled. A choice of units can make a single variable dominate the regression. Moreover, "standardization" begs the question of equivariance since it is just one of many ways to transform the variables into dimensionless numbers. Each such transformation produces a different orthogonal regression, and the relationships among the various regressions are not straightforward [Malinvaud (1966), chapter 1].

This lack of equivariance is evidently unsatisfactory. To some extent, it explains the popularity of ordinary least squares, where the regression coefficients adjust in an obvious and harmless way when any variable is rescaled [Morrison (1990), chapter 3].

We now propose a simple modification which makes orthogonal regression equivariant. This result is discussed in section 2, where a robust version is also described. Illustrative computations are provided in section 3.

2. A least-squares solution

Suppose that a data matrix X contains n joint observations on K variables ($n > K$). For convenience, all the variables are measured as deviations from their sample means. In the matrix equation

$$Xb = u \ , \tag{1}$$

b is a column vector of K regression coefficients and u is a column vector of n residuals. Orthogonal regression selects b to minimize the residual sum of squares

$$b'X'Xb = u'u \ . \tag{2}$$

A normalization is imposed to avoid the trivial solution $b = 0$. Conventionally, b is constrained to lie on the unit sphere:

$$b'b = 1 \ . \tag{3}$$

It then follows that b is the eigenvector corresponding to the smallest eigenvalue of X'X. However, we have emphasized that this solution lacks equivariance. Let us instead adopt the normalization

$$b'e = 1 \ , \tag{4}$$

where e is a column vector of K units. Accordingly, the sum of the regression coefficients is one. The Lagrangian expression

$$b'X'Xb - 2L(b'e - 1) \tag{5}$$

has a unique minimum at

$$L = 1/e'(X'X)^{-1}e \tag{6}$$

and $b = L(X'X)^{-1}e$. (7)

Equations (6) and (7) are the modified orthogonal regression which we propose; Raj [(1968), 16-17] has called this solution the "best weight function." Any computer software that handles matrices can easily calculate L and b. In fact, many statistical programs compute and display $(X'X)^{-1}$. We remark that the Lagrange multiplier L equals the minimum sum of squared residuals.

The coefficient vector b is equivariant in the following sense. Suppose that each observation on the first X variable is multiplied by a positive constant c. This rescaling means that the first row of X'X is multiplied by c; then the first column of X'X is multiplied by c. Consequently, the first row of $(X'X)^{-1}$ is multiplied by 1/c; then the first column of $(X'X)^{-1}$ is multiplied by 1/c.

If we now replace the first element of e by c, the normalization (4) becomes

$$cb_1 + b_2 + \ldots + b_K = 1 . \tag{8}$$

Then the rescaling has no effect on L in equation (6). In equation (7), b_1 is divided by c; but no other coefficient is altered. In summary, the rescaling affects our modified orthogonal regression just as it affects ordinary least squares.

Of course, it would usually be pointless to rescale an X variable and then nullify the effect by renormalizing, as in equation (8). Our intention is merely to show that the choice of units for an X variable is not a substantive decision, as indeed it should not be. We remark that Srinivasan (1976) uses a normalization like (8) in the context of ordinal regression.

If the X variables are not measured as deviations from the sample means, the model may require an intercept. It is computed as usual by passing the plane through the point of sample means [Malinvaud (1966), chapter 1].

When the X matrix may be contaminated by "outliers," a robust version of equations (6) and (7) can be calculated by the linear program

Maximize L subject to

$$\Sigma X_{ik} D_i + L = 0 \qquad \text{for } k = 1, \ldots, K \tag{9}$$

$$\text{and } -1 \le D_i \le 1 \qquad \text{for } i = 1, \ldots, n .$$

In (9), the summation over i runs from 1 to n. L is again the Lagrange multiplier for normalization (4). At the optimum, L equals the minimum sum of the absolute value of the residuals, $\Sigma|u_i|$. The residuals themselves are listed as "reduced costs." A variable $D_i = +1$ or -1 if the corresponding observation i lies above or below the regression plane; if the observation i lies right on the plane, then $-1 < D_i < 1$.

There are K constraints like (9), and the linear program reports a "dual variable" for each of them. These dual variables are the regression coefficients. To accommodate an intercept, the linear program may include constraint K+1: $\Sigma D_i = 0$. The solution by linear programming is related to (6) and (7) as a median is related to a mean, and this accounts for the robustness in the presence of outliers [Wagner (1959), Dodge (1987)].

3. Illustrative calculations

To illustrate equations (6), (7) and (8), we use some hypothetical data involving fifteen observations on five variables (n = 15, K = 5). The matrix X is:

1.2489	-1.2233	1.1348	-1.2265	.6205
.2365	.5172	.1794	.3618	.4656
-.3627	-.1500	-.2840	-.1923	-.3981
1.5916	1.8516	1.3597	1.7249	2.1422
.8176	.9119	.7665	.5742	1.0170
-2.3717	.1574	-2.1992	.0626	-2.0325
-.1758	.4104	-.2686	.1748	-.0274
-.2694	-2.1325	-.5668	-1.7472	-1.0196
.2092	1.1412	.5606	.9642	.5874
-.0537	1.8174	.2053	2.0861	.6575
-1.3818	-2.0502	-1.5132	-2.5885	-1.9695
-.1127	-1.2723	-.0885	-.9694	-.5930
-.6340	-1.0411	-.6080	-.5150	-.9685
.6595	.1888	.8501	.5806	.6420
.5984	.8735	.4719	.7098	.8765

So X'X =

13.8410	6.8952	13.2048	7.0396	14.8079
6.8952	23.0972	8.1500	21.9393	14.8312
13.2048	8.1500	13.0358	8.4323	14.6932
7.0396	21.9393	8.4323	22.0924	14.5169
14.8079	14.8312	14.6932	14.5169	18.6808

For $(X'X)^{-1}$ we have

109.3468	39.5861	-15.4423	2.7907	-108.1284
39.5861	15.8313	-4.1247	-.0300	-40.6805
-15.4423	-4.1247	5.1410	-.9884	12.2400
2.7907	-.0300	-.9884	1.0017	-2.1892
-108.1284	-40.6805	12.2400	-2.1892	110.1363

In equation (6), the Lagrange multiplier is the reciprocal of the sum of the elements of $(X'X)^{-1}$. For our example, L = 0.1329. In equation (7), b contains the five row sums of $(X'X)^{-1}$, each row sum having been multiplied by L:

$$b = (3.7420,\ 1.4065,\ -.4219,\ .0777,\ -3.8043)' \quad . \tag{10}$$

or $3.7420X_1 + 1.4065X_2 - .4219X_3 + .0777X_4 - 3.8043X_5 = 0$.

Of course, any variable may be expressed in terms of the others; for example:

$$X_2 = -2.6605X_1 + 0.3000X_3 - .0552X_4 + 2.7048X_5 \ .$$

To illustrate equivariance, we multiply each observation on the first variable by ten. The new X'X =

1384.0992	68.9524	132.0475	70.3956	148.0793
68.9524	23.0972	8.1500	21.9393	14.8312
132.0475	8.1500	13.0358	8.4323	14.6932
70.3956	21.9393	8.4323	22.0924	14.5169
148.0793	14.8312	14.6932	14.5169	18.6808

Accordingly, the new $(X'X)^{-1}$ is:

1.0935	3.9586	-1.5442	.2791	-10.8128
3.9586	15.8313	-4.1247	-.0300	-40.6805
-1.5442	-4.1247	5.1410	-.9884	12.2400
.2791	-.0300	-.9884	1.0017	-2.1892
-10.8128	-40.6805	12.2400	-2.1892	110.1363

In equations (6) and (7), we replace the unit vector e by (10, 1, 1, 1, 1)' and again obtain L = .1329. The regression coefficients are

$$b = (.3742,\ 1.4065,\ -.4219,\ .0777,\ -3.8043)' \quad . \qquad (11)$$

A comparison of (10) and (11) shows that the first coefficient has been divided by the scale factor of ten, but the other coefficients are unchanged. These results may also be compared with the coefficients in the usual orthogonal regression obtained from the smallest eigenvalue of X'X. Before the first variable is rescaled by ten, the eigenvector containing the regression coefficients is

$$(.6804,\quad .2518,\quad -.0868,\quad .0149,\quad -.6826)\ . \qquad (12)$$

After the first variable is rescaled by ten, the eigenvector is

$$(.0913,\quad .3445,\quad -.1056,\quad .0170,\quad -.9282)\ . \qquad (13)$$

The two eigenvectors, (12) and (13), are not related to one another by a straightforward transformation. On the other hand, the relationship between (10) and (11) is transparent.

The linear program for the robust orthogonal regression is shown below. An intercept (B0) has been included. The regression coefficients are not very different from (10), nor do there appear to be exceptionally large residuals in the column labeled REDUCED COST. It is therefore unlikely that the X matrix is contaminated by stray observations.

In conclusion, our modified orthogonal regression allows a researcher to explore a symmetric, equivariant and robust linear relationship among a set of variables.

Linear program for robust orthogonal regression

Maximize L subject to:

1.2489*D1+.2365*D2-.3627*D3+1.5916*D4+.8176*D5-2.3717*D6-.1758*D7-.2694*D8+.2092*D9-.0537*D10-1.3818*D11-.1127*D12-.634*D13+.6595*D14+.5984*D15+L=0

-1.2233*D1+.5172*D2-.15*D3+1.8516*D4+.9119*D5+.1574*D6+.4104*D7-2.1325*D8+1.1412*D9+1.8174*D10-2.0502*D11-1.2723*D12-1.0411*D13+.1888*D14+.8735*D15+L=0

1.1348*D1+.1794*D2-.284*D3+1.3597*D4+.7665*D5-2.1992*D6-.2686*D7-.5668*D8+.5606*D9+.2053*D10-1.5132*D11-.0885*D12-.608*D13+.8501*D14+.4719*D15+L=0

-1.2265*D1+.3618*D2-.1923*D3+1.7249*D4+.5742*D5+.0626*D6+.1748*D7-1.7472*D8+.9642*D9+2.0861*D10-2.5885*D11-.9694*D12-.515*D13+.5806*D14+.7098*D15+L=0

.6205*D1+.4656*D2-.3981*D3+2.1422*D4+1.017*D5-2.0325*D6-.0274*D7-1.0196*D8+.5874*D9+.6575*D10-1.9695*D11-.593*D12-.9685*D13+.642*D14+.8765*D15+L=0

(D1+..+D15)=0

1>=D1>=-1
1>=D2>=-1
1>=D3>=-1
1>=D4>=-1
1>=D5>=-1
1>=D6>=-1
1>=D7>=-1
1>=D8>=-1
1>=D9>=-1
1>=D10>=-1
1>=D11>=-1
1>=D12>=-1
1>=D13>=-1
1>=D14>=-1
1>=D15>=-1

VARIABLE		REDUCED COST
D1	-1.0000000	-.002865
D2	1.0000000	.201251
D3	-1.0000000	-.043831
D4	-.44177346	.000000
D5	1.0000000	-.191481
D6	-.69180276	.000000
D7	-1.0000000	-.142338
D8	1.0000000	.040054
D9	.69571016	.000000
D10	1.0000000	.030376
D11	1.0000000	.183603
D12	.66671341	.000000
D13	-1.0000000	-.068614
D14	-.22884735	.000000
D15	1.0000000	.002487
L	.90689890	.000000

B1	3.6788988
B2	1.3488624
B3	-.41810227
B4	.12282872
B5	-3.7324877
B0	-.00052750

References

Dodge, Y. (1987). Statistical Data Analysis Based on the L1-Norm and Related Methods. Amsterdam: North-Holland.

Malinvaud, E. (1966). Statistical Methods of Econometrics. Chicago: Rand McNally.

Morrison, D. F. (1990). Multivariate Statistical Methods. New York: McGraw-Hill.

Raj, D. (1968). Sampling Theory. New York: McGraw-Hill.

Srinivasan, V. (1976). "Linear Programming Computational Procedures for Ordinal Regression," Journal of the Association for Computing Machinery, 23 (3), 475-487.

Wagner, H. (1959). "Linear Programming Techniques for Regression Analysis," Journal of the American Statistical Association, 56, 206-212.